

IN THE CLAIMS

Please amend the claims as follows:

Claim 1 (Currently Amended): A computer processing apparatus for classifying a document, comprising:

B1  
~~means for accessing~~ a database having a database structure providing a classification scheme having a plurality of different subject matter categories, the database containing a classified vocabulary including a plurality of terms in each of the different subject matter categories with each term being classified in accordance with the ~~subject matter category structure of the database~~ classification scheme and the database also containing a classification data set comprising a plurality of groups of terms with each group being associated with a specific different one of the subject matter categories and each group including a plurality of terms exemplifying the associated category for facilitating disambiguation between different meanings of the same term;

means for receiving in computer-readable form a ~~text~~ document to be classified;

processor means ~~operable to compare~~ for comparing terms appearing in the text document with the terms ~~in the classified vocabulary in the database~~ and ~~to determine~~ for determining from the comparison the category for the document; and

means for supplying a signal carrying data representing the ~~text~~ document and data associating the ~~text~~ document with the determined category.

Claims 2-3 (Canceled).

B2  
Claim 4 (Currently Amended): The apparatus according to claim 1, wherein the processor means ~~is operable to determine~~ determines the category for the document by determining from the comparison the category or categories of terms in the document,

B2 ~~assigning assigns~~ weightings to the determined categories for the terms, and ~~assigning assigns~~ the document being classified to the category having the highest weighting.

---

Claims 5-6 (Canceled).

---

B3 Claim 7 (Currently Amended): The apparatus according to claim 4, wherein the processor means ~~is operable~~ shares, for each term in the classified vocabulary and in the text document, ~~to share~~ a predetermined weighting factor between each category associated with the term.

---

Claims 8-11 (Canceled).

---

Claim 12 (Currently Amended): A computer processing apparatus for classifying a document, comprising:

B4 means for accessing a database having a database structure providing a plurality of different subject matter categories, the database containing a classified vocabulary including a plurality of terms in each of the different subject matter categories with each term being classified in accordance with the subject matter category structure of the database and the database also containing a plurality of collocations each collocation being associated with a specific different one of the subject matter categories and each collocation including a plurality of terms exemplifying the associated category for disambiguating a different meaning of the same term;

means for receiving in computer-readable form a text document to be classified;

processor means ~~operable to compare~~ for comparing terms appearing in the text document with the collocations to determine the collocation having the most terms in

common with the document, and ~~to allocate~~ for allocating the category of the determined collocation to the document; and

B4 means for supplying a signal carrying data representing the text document and data associating the text document with the determined category.

---

Claims 13-16 (Canceled).

---

B5 Claim 17 (Currently Amended): ~~Apparatus~~ The apparatus according to claim 12, wherein the processor means ~~is operable to disambiguate~~ disambiguates between different meanings of terms by using the collocations.

---

Claim 18 (Canceled).

---

B6 Claim 19 (Currently Amended): The apparatus according to claim ~~[[7]]~~ 12, wherein the accessing means ~~is arranged to access~~ accesses the collocations from store means separate from the remainder of the database.

---

Claim 20 (Canceled).

---

Claim 21 (Currently Amended): The apparatus according to claim 1, further comprising ~~store~~ means for storing the database.

---

B7 Claim 22 (Currently Amended): The apparatus according to claim 1, wherein the database ~~structure~~ provides said plurality of subject matter categories as a tree structure

including a plurality of main subject matter areas each divided into two or more subject matter areas.

Claim 23 (Currently Amended): The apparatus according to claim 1, wherein the database ~~structure~~ provides said plurality of subject matter categories such that each category is defined by a subject matter area and a species or genus.

Claim 24 (Previously Presented): The apparatus according to claim 23, wherein the database provides said plurality of subject matter categories such that the species or ~~geni~~ are people, places, organisations, products and technology.

Claim 25 (Currently Amended): The apparatus according to claim 23, wherein the database ~~structure~~ provides said plurality of subject matter categories such that the species or ~~genus~~ geni are the same for each subject matter area.

Claim 26 (Previously Presented): The apparatus according to claim 1, wherein the database provides categories in each of the following subject matter areas: the universe, the earth, the environment, natural history, humanity, recreation, society, the mind and human history.

Claim 27 (Currently Amended): The apparatus according to claim 23, wherein the database ~~structure~~ is such that, for a given meaning, a term is associated with only one category and different meanings of the same term are associated with different categories.

Claim 28 (Previously Presented): The apparatus according to claim 1, wherein the supplying means comprises means for storing a signal supplied by the supplying means on a computer readable medium.

Claim 29 (Previously Presented): The apparatus according to claim 1, wherein the supplying means comprises means for forwarding a signal supplied by the supplying means to another processing apparatus.

Claim 30 (Previously Presented): The apparatus according to claim 1, wherein the supplying means comprises means for displaying the information to a user.

B7  
Claim 31 (Currently Amended) In a computer processing apparatus having ~~means for~~ accessing a database having a database structure providing a classification scheme having a plurality of different subject matter categories, the database containing a classified vocabulary including a plurality of terms in each of the different subject matter categories with each term being classified in accordance with the ~~subject-matter category structure of the database~~ classification scheme and the database also containing a classification data set comprising a plurality of groups of terms with each group being associated with a specific different one of the subject matter categories and each group including a plurality of terms exemplifying the associated category whereby the classification data set facilitates disambiguation between different meanings of the same term, and ~~means for receiving a receiver configured to receive~~ in computer-readable form a text document to be classified, a method of classifying documents comprising:

comparing terms appearing in the text document with the terms in the ~~classified~~ vocabulary database;

B7  
determining from the comparison the category for the text document; and  
supplying a signal carrying data representing the text document and data associating  
the text document with the determined category.

---

Claims 32-33 (Canceled).

---

B8  
Claim 34 (Previously Presented): The method according to claim 31, further  
comprising determining the category for the document by determining from the comparison  
the category or categories of the terms in the document, assigning weightings to the  
determined categories for the terms, and assigning the document being classified to the  
category having the highest weighting.

Claim 35 (Previously Presented): The method according to claim 34, further  
comprising assigning weighting by, for each term in the classified vocabulary and in the text  
document, sharing a predetermined weighting factor between each category associated with  
the term.

---

Claims 36-38 (Canceled).

---

B9  
Claim 39 (Currently Amended): In a computer processing apparatus having ~~means~~  
~~for accessing~~ a database having a database structure providing a classification scheme having  
a plurality of different subject matter categories, the database containing a classified  
vocabulary including a plurality of terms in each of the different subject matter categories  
with each term being classified in accordance with the ~~subject-matter category structure of the~~  
~~database and the database also containing a plurality of collocations each collocation being~~

39 associated with a specific different one of the subject matter categories and each collocation including a plurality of terms exemplifying the associated category and having classification scheme and the database also containing a classification data set comprising a plurality of collocations of terms with each collocation being associated with a specific different one of the subject matter categories and each collocation including a plurality of terms exemplifying the associated category for disambiguating different meanings of the same term, and means for receiving a receiver configured to receive in computer-readable form a text document to be classified, a method of classifying documents comprising:

comparing terms appearing in the text document with the collocations to determine the collocation having the most terms in common with the text document;

allocating the category of the determined collocation to the document; and

supplying a signal carrying data representing the text document and data associating the text document with the determined category.

---

Claims 40-42 (Canceled).

310 Claim 43 (Currently Amended): The method according to claim 36 39, further comprising accessing the collocations from store means separate from the remainder of the database.

---

Claims 44-49 (Canceled).

311 Claim 50 (Previously Presented): The method according to claim 31, further comprising carrying out the supplying by storing a signal on a computer-readable medium.

Claim 51 (Previously Presented): The method according to claim 31, further comprising carrying out the supplying by forwarding a signal to another processing apparatus.

Claim 52 (Previously Presented): The method according to claim 31, further comprising displaying the information to a user.

B11  
Claim 53 (Currently Amended): A database for use with an apparatus in accordance with claim 1, the database having a database structure providing a classification scheme having a plurality of different subject matter categories, the database containing a classified vocabulary including a plurality of terms in each of the different subject matter categories with each term being classified in accordance with the ~~subject-matter category structure of the database~~ classification scheme and the database also containing a classification data set comprising a plurality of groups of terms with each group being associated with a specific different one of the subject matter categories and each group including a plurality of terms exemplifying the associated category whereby the classification data set facilitates disambiguation between different meanings of the same term.

---

Claims 54-55 (Canceled).

---

B12  
Claim 56 (Currently Amended): A database for use with an apparatus in accordance with claim 12, the database having a database structure providing a classification scheme having a plurality of different subject matter categories, the database containing a classified vocabulary including a plurality of terms in each of the different subject matter categories with each term being classified in accordance with the ~~subject-matter category structure of the~~



B12  
database classification scheme and the database also containing a classification data set comprising a plurality of collocations each collocation being associated with a specific different one of the subject matter categories and each collocation including a plurality of terms exemplifying the associated category whereby the classification data set facilitates disambiguation between different meanings of the same term.

---

Claims 57-64 (Canceled).

---

Claim 65 (Currently Amended): An apparatus for classifying electronic documents, comprising:

storage means for storing a classification scheme having a plurality of collocations, each collocation being associated with a respective different subject matter area and containing a set of terms which exemplify that subject matter area and for facilitating disambiguation between different meanings of the same term;

B13  
means for comparing terms used in a document to be classified with the terms in said collocations;

means for allocating the document being classified to the one of said collocations which said comparing means identifies as having the most number of terms in common with the document being classified;

means for associating with the document being classified a code representing the subject matter area of the ~~allocation~~ allocated collocation; and

means for storing the document together with the associated code.

---

Claims 66-69 (Canceled).

B14  
Claim 70 (Original): A signal carrying processor implementable instructions for causing apparatus to become configured to form apparatus in accordance with claim 1.

Claims 71-72 (Canceled).

Claim 73 (Currently Amended): A signal carrying a database in accordance with claim 53 ~~or a plurality of collocations for use with the database.~~

Claim 74 (Currently Amended): A storage medium carrying a database in accordance with claim 53 ~~or a plurality of collocations for use with the database.~~

Claim 75 (Currently Amended): A processor readable medium storing processor readable instructions for causing a processor to:

B15  
access a database having a database structure providing a classification scheme having a plurality of different subject matter categories, the database containing a classified vocabulary consisting of including a plurality of terms in all each of the different subject matter categories with each term being classified in accordance with the subject matter category structure of the database classification scheme and the database also containing a classification data set comprising a plurality of groups of terms with each group being associated with a specific different one of the subject matter categories and each group including a plurality of terms exemplifying the associated category for facilitating disambiguation of different meanings of the same term; and means for receiving receive in computer-readable form a text document to be classified;  
compare terms appearing in the text document with the terms in the classified vocabulary database;

B15  
determine from the comparison the category for the document; and  
supply a signal carrying data representing the text document and data associating the  
text document with the determined category.

---

Claims 76-78 (Canceled).

---

Claim 79 (Currently Amended): A computer processing apparatus for classifying  
documents, ~~[[the-apparatus]]~~ the apparatus comprising:

B16  
a database having a database structure defining a classification scheme for terms,  
the classification scheme having subject matter data defining main and subsidiary  
subject matter domains into which terms can be classified and genera data defining a  
predetermined number of genera to which terms can be allocated, the classification scheme  
being such that a term can be allocated to more than one subject matter domain but to only  
one genus so that each specific combination of subsidiary subject matter domain and genus  
defines a unique category,

the database also having classified vocabulary comprising a set of terms classified in  
accordance with the classification scheme such that each term is associated with category  
data identifying the corresponding category,

the database also including a classification scheme data set which includes a  
respective different classification scheme data set item associated with each category,

each classification scheme data set item comprising a collocation consisting of a list  
of terms that may be used to describe the function, appearance or relationship with other  
objects of classified terms in that category or that may be used in relation to terms in that  
category,

a receiver operable to receive in computer-readable form a text document to be classified;

a processor configured to compare terms in the text document with terms in at least one of the classified vocabulary and the collocations to determine a category for the text document; and

a signal supplier configured to supply a signal carrying data representing the text document and data associating the text document with the determined category data.

Claim 80 (Previously Presented): A method of classifying documents, the method comprising:

providing a classification scheme having subject matter data defining main and subsidiary subject matter domains into which terms can be classified and genera data defining a predetermined number of genera to which terms can be allocated, the classification scheme being such that a term can be allocated to more than one subject matter domain but to only one genus so that each specific combination of subsidiary subject matter domain and genus defines a unique category;

providing a classified vocabulary comprising a set of terms classified in accordance with the classification scheme such that each term in the classified vocabulary is associated with category data identifying the corresponding category;

providing a classification scheme data set which includes a respective different classification scheme data set item associated with each category with each classification scheme data set item comprising a collocation consisting of a list of terms that may be used to describe the function, appearance or relationship with other objects of classified terms in that category or that may be used in relation to terms in that category;

receiving data representing a text document to be classified; and

B16 comparing terms in the text document with terms in at least one of the classified vocabulary and the collocations to determine a category for the text document.

---

Claim 81 (New): A computer processing apparatus for classifying documents, the apparatus comprising:

B17 a database having a database structure providing a classification scheme having a plurality of different subject matter categories, the database containing a classified vocabulary consisting of a plurality of terms in each of the different subject matter categories with each term being classified in accordance with the classification scheme and the database also containing a classification data set comprising a plurality of groups of terms with each group being associated with a specific different one of the subject matter categories and each group including terms that may be used to describe the function, appearance or relationship with other objects of classified terms in that category or that may be used in relation to terms in that category to facilitate disambiguation between different meanings of the same term;

a receiver configured to receive in computer-readable form a text document to be classified;

a processor configured to use the groups of terms in the classification data set to disambiguate different meanings of terms in the document and to determine a category for the text document using the database; and

a signal supplier configured to supply a signal carrying data representing the text document and data associating the text document with the determined category data.

---